

Claim 1. A method for modifying an audio visual recording originally produced with an original audio track of an original speaker, using a second audio dub track of a second speaker, to produce a new audio visual recording with synchronized audio to facial expressive speech of the second audio dub track spoken by the original speaker,

5 comprising analyzing the original audio track to convert it into phonemes as a time-coded phoneme stream to identify corresponding visual facial motions of the original speaker to create frames of facial motion corresponding to speech phoneme utterance states and transformations, storing these frames in a database, analyzing the second audio dub track to convert it to phonemes as a time-coded phoneme stream, using the second audio dub

10 track time-coded phoneme stream to animate the original speaker's face, synchronized to the second audio dub track to create natural continuous facial speech expression by the original speaker of the second dub audio track.

Claim 2. The method of Claim 1 wherein said second audio dub track is spoken in a

15 language different from that of the original speaker.

Claim 3. The method of Claim 1 wherein said phonemes comprise diphones.

Claim 4. The method of Claim 1 wherein said phonemes comprise triphones.

20

Claim 5. The method of Claim 1 comprising using a set of fixed facial reference points to track the facial transformation from one phoneme to another phoneme.

Claim 6. The method of Claim 5 comprising using a computer motion tracking system to record the optical flow path for each fixed reference point.

Claim 7. The method of Claim 1 further comprising using a set of fixed facial reference  
5 points to track the facial transformation from one phoneme to another phoneme.

Claim 8. The method of Claim 7 further comprising accumulating a database of recorded triphone and diphone mouth transformation optical flow paths.

10 Claim 9. The method of Claim 5 further comprising adding an emotional elicitation process by using a computer motion tracking system to record a number of facial control points for each emotion.

15 Claim 10. The method of Claim 9 in which the facial control points comprise the chin, outside of the mouth and inside of the lips.

Claim 11. The method of Claim 7 comprising accumulating a database of visemes by recording the facial control points corresponding to each phoneme.

20 Claim 12. The method of Claim 11 comprising accumulating a database of muzzle patches by mapping the facial control points for each viseme.

Claim 13. The method of Claim 12 comprising selecting muzzle patches from the speaker's viseme database based on the second dub audio track and phoneme to viseme sequence and applying the selected muzzle patches onto a three dimensional facial muzzle model.

5

Claim 14. The method of Claim 13 further comprising also collecting a set of patches of light, color and texture from the sampled first speaker's muzzle patches.

Claim 15. The method of Claim 1 in which the fixed reference points for both speakers  
10 are mapped using radar.

Claim 16. The method of Claim 16 in which the dub audio track radar measurements are referenced to a particular phoneme or phoneme to phoneme transition in time.

15 Claim 17. A method for modifying an audio visual recording originally produced with an original audio track of an original screen actor, using a second audio dub track of a second screen actor, to produce a new audio visual recording with synchronized audio to facial expressive speech of the second audio dub track spoken by the original screen actor, comprising analyzing the original audio track to convert it into phonemes as a  
20 time-coded phoneme stream to identify corresponding visual facial motions of the original speaker to create frames of facial motion corresponding to speech phoneme utterance states and transformations, storing these frames in a database, analyzing the second audio dub track to convert it to phonemes as a time-coded phoneme stream, using

the second audio dub track time-coded phoneme stream to animate the original screen actor's face, synchronized to the second audio dub track to create natural continuous facial speech expression by the original screen actor of the second dub audio track.

5 Claim 18. The method of Claim 17 in which both audio tracks are time stamped to frames to create a database of individual frames for each phoneme.

Claim 19. The method of Claim 18 comprising using a computer vision system to track and record a database of visemes of the actors' head position, facial motions of the jaw, 10 and the lip motion during speech for each frame.

Claim 20. The method of Claim 19 comprising tracking fixed reference control points on the head, jaw and lips.

15 Claim 21. The method of Claim 20 in which the reference control points comprise the outside edge of the lips, the inside edge of the lips, the edge of the chin, the tip of the tongue, the bottom edge of the upper teeth, the upper edge of the lower teeth, the nose and the eyes.

20 Claim 22. The method of Claim 21 comprising accumulating a database of muzzle patches by mapping the facial control points for each viseme.

Claim 23. The method of Claim 22 comprising selecting muzzle patches from the speaker's viseme database based on the second dub audio track and phoneme to viseme sequence and applying the selected muzzle patches onto a three dimensional facial muzzle model.

5

Claim 24. The method of Claim 23 further comprising also collecting a set of patches of light, color and texture from the sampled first speaker's muzzle patches and applying these patches based on the second dub audio track.

TECHNICAL FIELD

10 Claim 25. The method of Claim 20 in which the fixed reference points for both actors are mapped using radar.

Claim 26. The method of Claim 25 in which the dub audio track radar measurements are referenced to a particular phoneme or phoneme to phoneme transition in time.

15

Claim 27. The method of Claim 25 in which the radar mapping information is transferred to an animation control mixer.

20 Claim 28. The method of Claim 25 in which the viseme database information is transferred to an animation control mixer.

Claim 29. The methods of Claim 1 or 17 comprising using head modeling to aid in auto-positioning three dimensional visemes.

Claim 30. The methods of Claim 1 or 17 comprising modeling multiple floating point control vertices relative to a generic head shape and relative to different speech viseme standards.

5     Claim 31. The methods of Claim 1 or 17 comprising texture sampling target footage to texture match visemes.

Claim 32. A method for modifying an audio visual recording originally produced with an original audio track of an original screen actor, using a second audio dub track of a second screen actor, to produce a new audio visual recording with synchronized audio to facial expressive speech of the second audio dub track spoken by the original screen actor, comprising analyzing the original audio track to convert it into phonemes as a time-coded phoneme stream, identifying corresponding visemes of the original screen actor, using radar to measure a set of facial reference points corresponding to speech phoneme utterance states and transformations, storing the data obtained in a database, analyzing the second audio dub track to convert it to phonemes as a time-coded phoneme stream, identifying corresponding visemes of the second screen actor, using radar to measure a set of facial reference points corresponding to speech phoneme utterance states and transformations, storing the data obtained in a database, using the second audio dub track time-coded phoneme stream and the actors visemes to animate the original screen actor's face, synchronized to the second audio dub track to create natural continuous facial speech expression by the original screen actor of the second dub audio track.

Claim 33. The method of Claim 32 in which the database resides in an animation control mixer.

Claim 34. The method of Claim 32 in which the facial reference points comprise the

5 outside edge of the lips, the inside edge of the lips, the edge of the chin, the tip of the tongue, the bottom edge of the upper teeth, the upper edge of the lower teeth, the nose and the eyes.

Claim 35. The method of Claim 32 comprising texture sampling target footage to texture

10 match visemes.